確率・統計基礎: 中心極限定理、正規分布

森 立平

大数の法則と集中不等式

このスライドでは $X_1, \ldots, X_N$ は i.i.d. で確率変数 X と同分布であると仮定する。

集中不等式は

$$\Pr\left(\left|\left(\sum_{k=1}^{N} X_k\right) - N\mathbb{E}\left[X\right]\right| \geq N\epsilon\right)$$

が指数関数的に小さいことを示している。

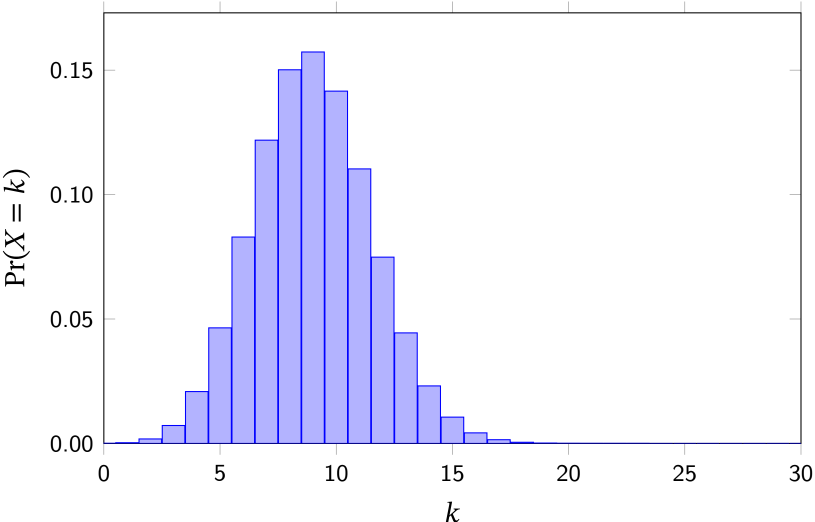
今日は期待値の近くに注目し、

$$\Pr\left(\left|\left(\sum_{k=1}^{N} X_k\right) - N\mathbb{E}\left[X\right]\right| \geq \sqrt{N}\epsilon\right)$$

について考えてみる

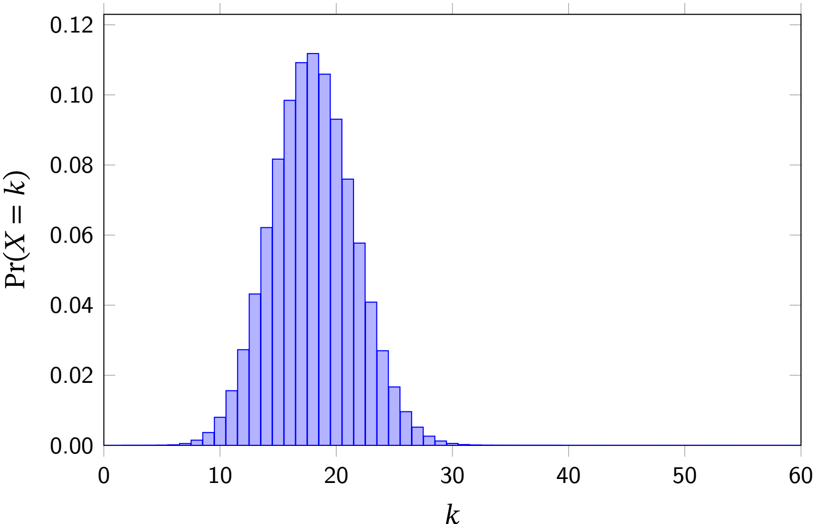
二項分布

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad n = 30,\ p = 0.3.$$



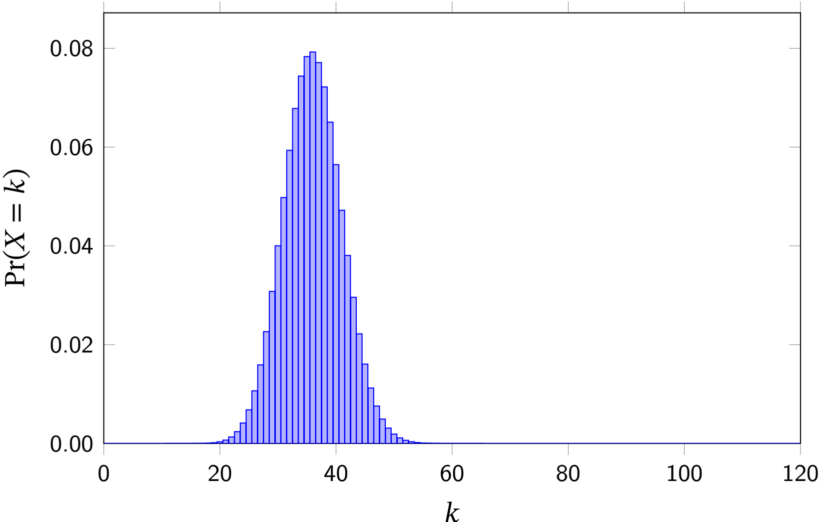
二項分布

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad n = 60,\ p = 0.3.$$



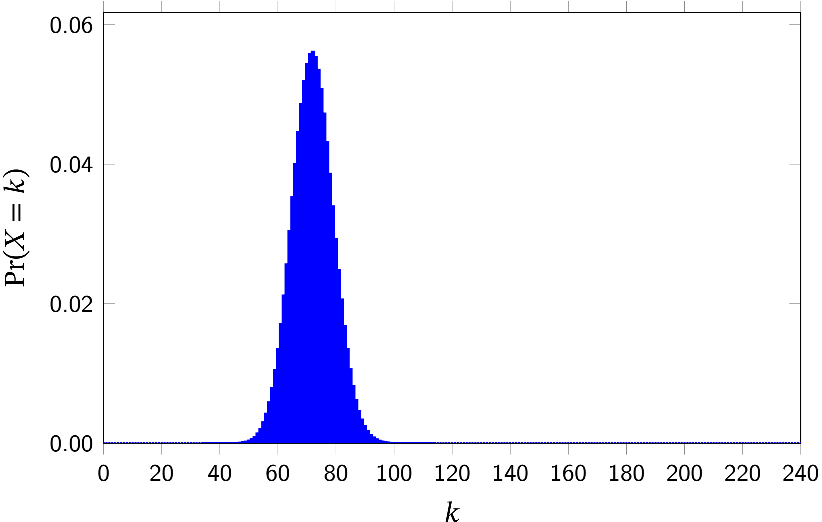
二項分布

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad n = 120,\ p = 0.3.$$



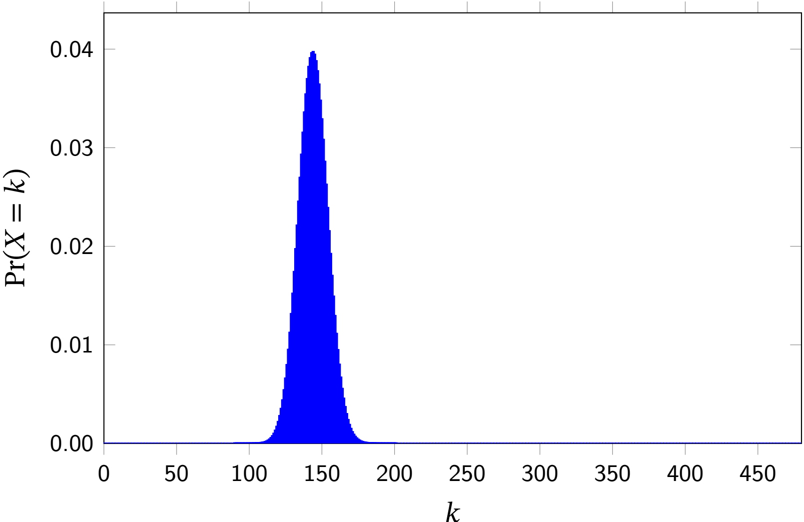
二項分布

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad n = 240,\ p = 0.3.$$



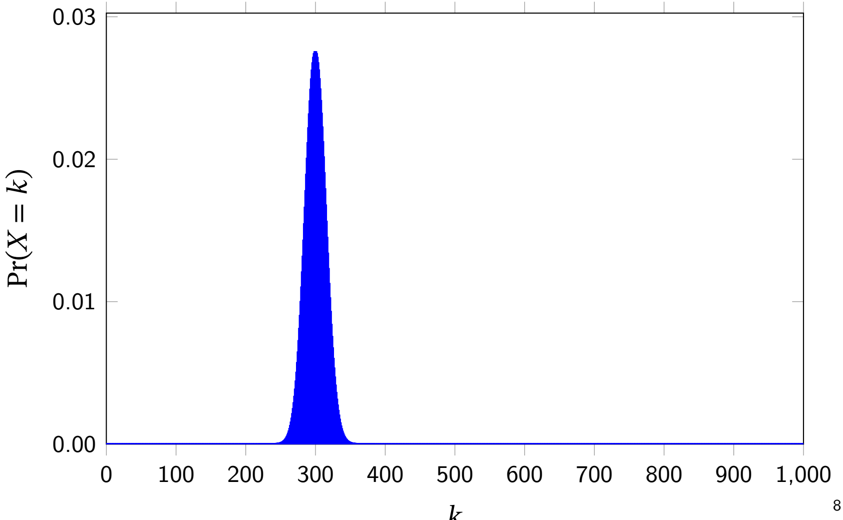
二項分布

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad {\color{red} n = 480}, \ p = 0.3.$$



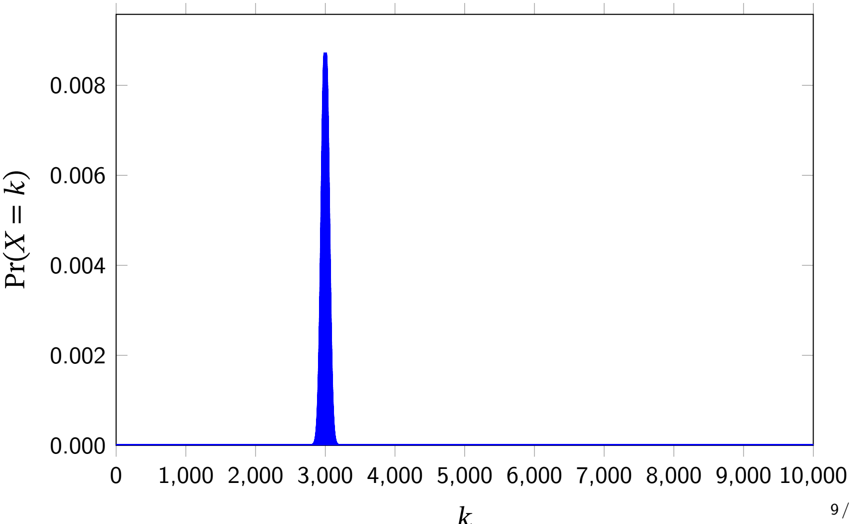
二項分布

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad n = 1000,\ p = 0.3.$$



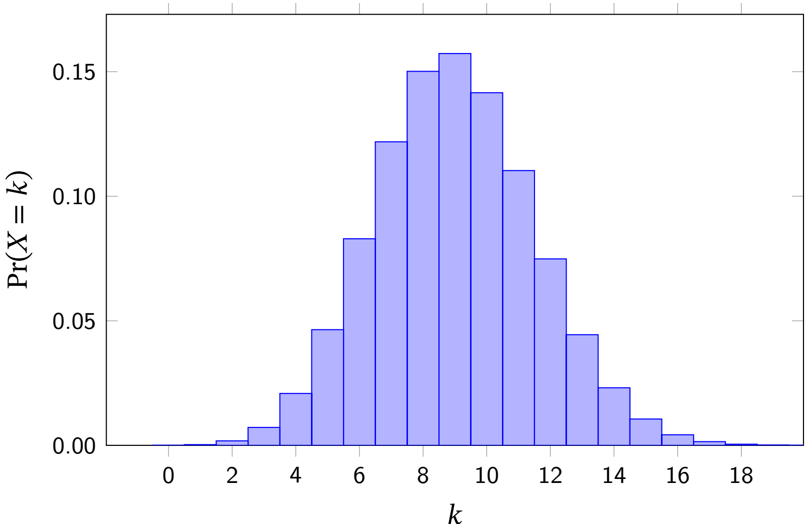
二項分布

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad {\color{red} n = 10000}, \ p = 0.3.$$



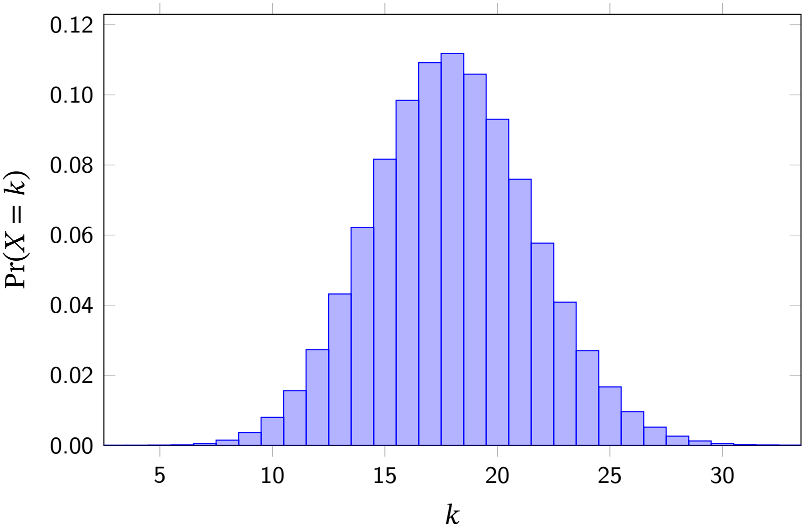
二項分布

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad n = 30,\ p = 0.3.$$



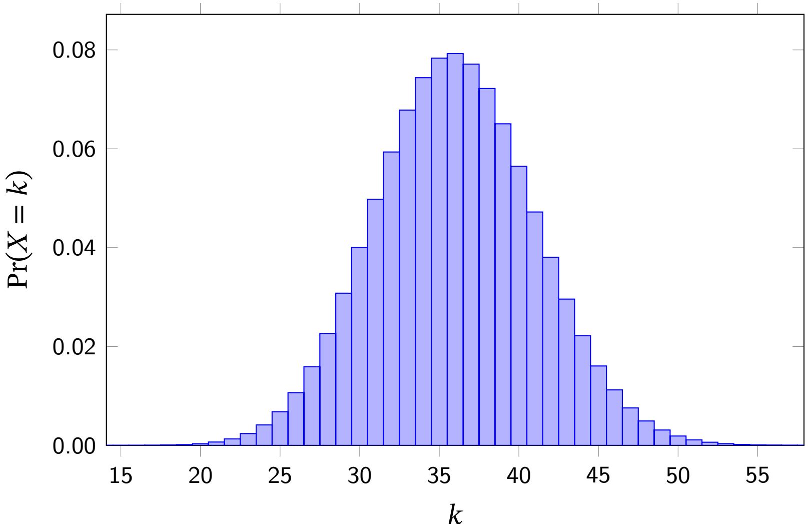
二項分布

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad n = 60,\ p = 0.3.$$



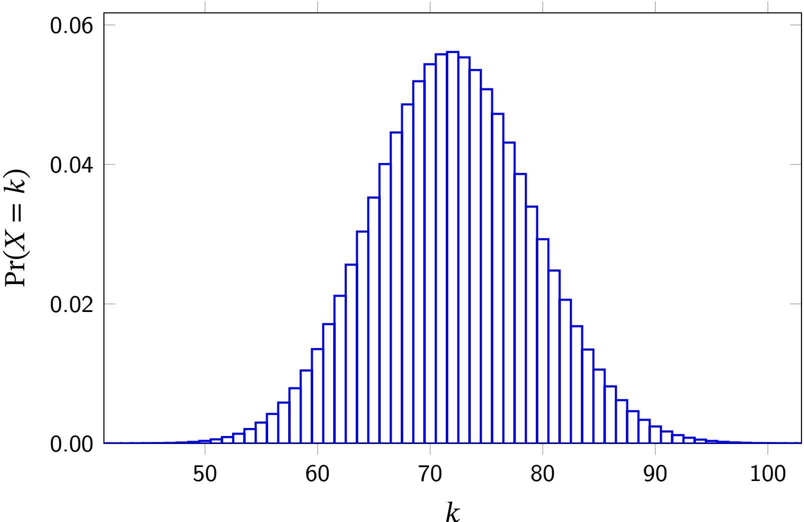
二項分布

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad n = 120,\ p = 0.3.$$



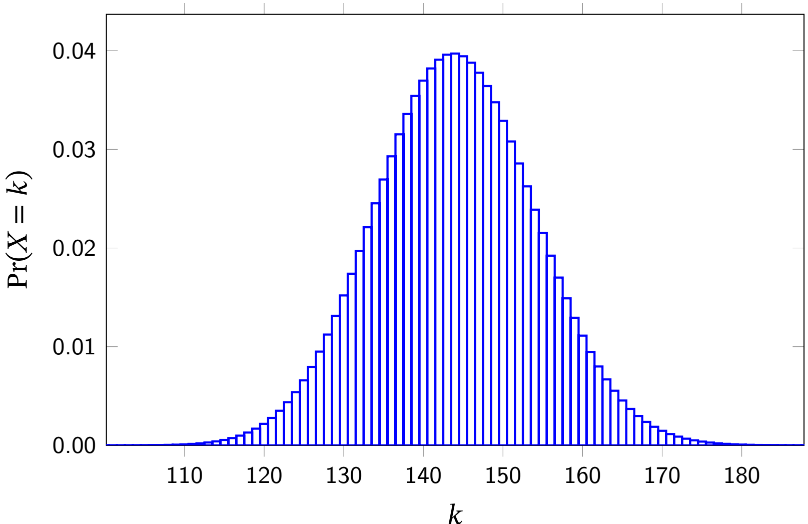
二項分布

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad n = 240,\ p = 0.3.$$



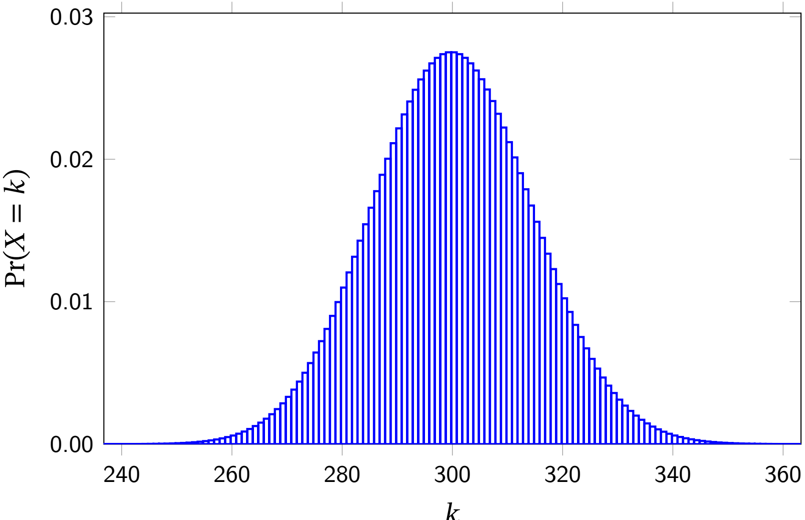
二項分布

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad n = 480,\ p = 0.3.$$



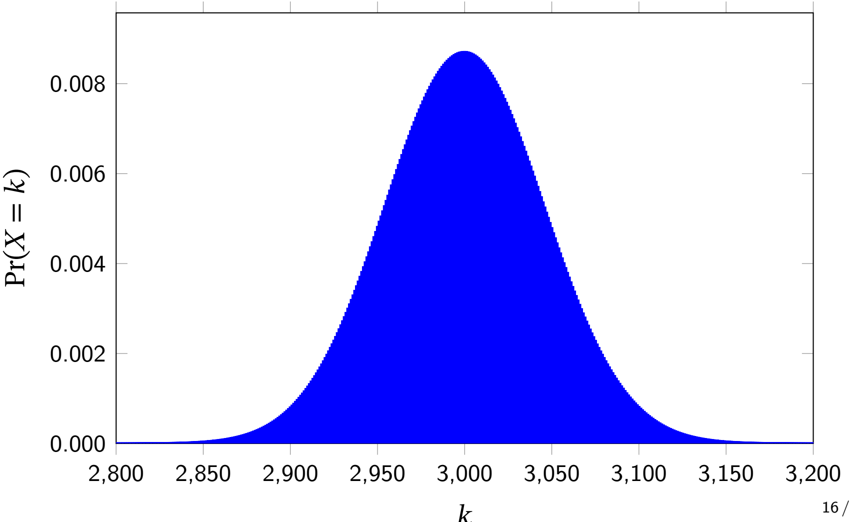
二項分布

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad n = 1000, \; p = 0.3.$$



二項分布

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad n = 10000,\ p = 0.3.$$



中心極限定理

Theorem (中心極限定理)

X が分散を持つと仮定する。

$$\lim_{N\to\infty} \Pr\left(\frac{\left(\sum_{k=1}^{N} X_k\right) - N\mathbb{E}\left[X\right]}{\sqrt{\mathbb{V}\left[X\right]N}} \le x\right) = \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \mathrm{e}^{-\frac{t^2}{2}}\,\mathrm{d}t$$

$$Y_k := \frac{X_k - \mathbb{E}\left[X\right]}{\sqrt{\mathbb{V}\left[X\right]}} \qquad \text{for } k = 1, 2, \dots, N$$

は X_k を期待値 0、分散 1 に正規化した確率変数である。中心極限定理の主張は

$$\lim_{N\to\infty} \Pr\left(\frac{\sum_{k=1}^{N} Y_k}{\sqrt{N}} \le x\right) = \Phi(x)$$

である。N で割るのではなくて $\sqrt{N}$ で割ることで大数の法則よりも細かい部分を見ている。

特性関数

Definition (特性関数)

確率変数 X の特性関数 $\varphi_X\colon \mathbb{R} \to \mathbb{C}$ を以下で定義する。

$$\varphi_X(t) := \mathbb{E}\left[\mathrm{e}^{itX}\right].$$

モーメント母関数は存在するとは限らないが特性関数は (定義域を $\mathbb{R}$ として) 常に存在する。

X と Y の分布が同じ $\Longleftrightarrow$ $\varphi_X = \varphi_Y$ (ルベーグ積分に基づいた議論が必要).

X が k 次モーメントを持つ $\Longleftrightarrow$ φ_X は 0 で k 回微分できる。

$$\left.\frac{\mathrm{d}^k \varphi_X(t)}{\mathrm{d}t^k}\right|_{t=0} = i^k \mathbb{E}\left[X^k\right].$$

XとYの分布が同じ $\iff$ $\varphi_X = \varphi_Y$ (像が有限の場合)

確率変数 X の像 $\mathrm{Image}(X)$ が有限集合のとき、

$$\varphi_X(t) = \sum_{k \in \mathrm{Image}(X)} \Pr(X = k) \mathrm{e}^{itk}$$

ここで、$\mathrm{Image}(X)$ の最小値と最大値をそれぞれ k_{min} と k_{max} とおくと、$k_{\mathsf{max}} - k_{\mathsf{min}}$ 次多項式 Q_X を以下のように定義する。

$$Q_X(x) = \sum_{k \in \mathrm{Image}(X)} \Pr(X = k) x^{k - k_{\mathsf{min}}}$$

このとき、$Q_X(\mathrm{e}^{it}) = \varphi_X(t)\mathrm{e}^{-itk_{\mathsf{min}}}$ である。一般的に n 次多項式は異なる $(n+1)$ 点での値から一意に定まる。よって特性関数から多項式 Q_X が一意に定まり、$\Pr(X = k)$ が一意に定まる。

確率質量関数から特性関数への写像が単射であることが分かった。これは一般の確率変数について (確率密度関数が存在しない場合でも) 成り立つ。

おまけ: N 次多項式は $N+1$ 点での値から定まる

Lemma

高々 N 次の体上の多項式 P と Q が $N+1$ 個の異なる点で同じ値を取るとき、$P = Q$ である。

Proof.

仮定より、$P(x) - Q(x)$ は高々 N 次の多項式であり、$N+1$ 個の異なる点で 0 である。一般に高々 N 次の非ゼロ多項式は高々 N 個の異なる根を持つ。そのため $P - Q$ はゼロ多項式である。よって $P = Q$ である。 □

分布の収束と特性関数の収束

Theorem (レヴィの連続性定理)

確率変数の列 $Z_1, Z_2, \ldots$ と確率変数 Z について

$$\lim_{n\to\infty} \varphi_{Z_n}(t) = \varphi_Z(t) \qquad \forall t \in \mathbb{R}$$

のとき、

$$\lim_{n\to\infty} \Pr(Z_n \le x) = \Pr(Z \le x) \qquad \forall x \in \mathbb{R}$$

分布の収束を示すには特性関数の収束を示せばよい。

中心極限定理の証明

期待値 0 分散 1 の i.i.d. 確率変数 $Y_1, \dots, Y_N$ と標準正規分布に従う確率変数 Z について

$$\lim_{N\to\infty} \Pr\left(\frac{\sum_{k=1}^N Y_k}{\sqrt{N}} \le x\right) = \Pr(Z \le x)$$

を示すためには

$$\lim_{N\to\infty} \varphi_{\frac{\sum_k Y_k}{\sqrt{N}}}(t) = \varphi_Z(t) = \mathrm{e}^{-\frac{t^2}{2}}$$

を示せばよい。Y は分散を持つので φ_Y は 0 で二回微分できる。テイラーの定理 (φ_Y は $\mathbb{R} \to \mathbb{C}$ の関数だけど) より

$$\begin{aligned}
\varphi_{\frac{\sum_k Y_k}{\sqrt{N}}}(t) &= \varphi_{\sum_k Y_k}\left(\frac{t}{\sqrt{N}}\right) = \varphi_Y\left(\frac{t}{\sqrt{N}}\right)^N \\
&= \left(\varphi_Y(0) + \varphi_Y'(0)\frac{t}{\sqrt{N}} + \frac{\varphi_Y''(0)}{2}\left(\frac{t}{\sqrt{N}}\right)^2 + o\left(\frac{1}{N}\right)\right)^N \\
&= \left(1 - \frac{1}{2}\frac{t^2}{N} + o\left(\frac{1}{N}\right)\right)^N \xrightarrow{N\to\infty} \mathrm{e}^{-\frac{t^2}{2}}
\end{aligned}$$

中心極限定理の意義

たくさんの独立同分布の確率変数の和は期待値の近くでは正規分布のように振る舞う。

同分布という条件は緩められる。

多くの不確かな要因からくるノイズは正規分布とみなすとよい近似になっている。

正規分布の性質

平均 μ、分散 σ^2 の正規分布の確率密度関数は

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

X が平均 μ、分散 σ^2 の正規分布に従うことを、$X \sim N(\mu, \sigma^2)$ と表す。

- 任意の $a \in \mathbb{R}$ について $X \sim N(\mu, \sigma^2)$ のとき
  $X + a \sim N(\mu + a, \sigma^2)$ である。

- 任意の $a \in \mathbb{R} \setminus \{0\}$ について $X \sim N(\mu, \sigma^2)$ のとき
  $aX \sim N(a\mu, a^2\sigma^2)$ である。

- 独立な確率変数 $X_1 \sim N(\mu_1, \sigma_1^2)$ と $X_2 \sim N(\mu_2, \sigma_2^2)$ について
  $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ である。

$X + a$ の確率密度関数

確率変数 X が確率密度関数 $p(x)$ を持つとする。
このとき、$a \in \mathbb{R}$ について確率変数 $X + a$ の確率密度関数は

$$\begin{aligned}
\Pr(X + a \leq x) &= \Pr(X \leq x - a) \\
&= \int_{-\infty}^{x-a} p(t)\mathrm{d}t \\
&= \int_{-\infty}^{x} p(u - a)\mathrm{d}u \qquad (u = t + a)
\end{aligned}$$

より $p(x - a)$ である。

aX の確率密度関数

また、$a > 0$ について確率変数 aX の確率密度関数は

$$
\begin{aligned}
\Pr(aX \leq x) &= \Pr\left(X \leq \frac{x}{a}\right) \\
&= \int_{-\infty}^{\frac{x}{a}} p(t)\mathrm{d}t \\
&= \int_{-\infty}^{x} p\left(\frac{u}{a}\right)\frac{1}{a}\mathrm{d}u \qquad (u = ta)
\end{aligned}
$$

より $\frac{1}{a}p\left(\frac{x}{a}\right)$ である。また、$a < 0$ について確率変数 aX の確率密度関数は

$$
\begin{aligned}
\Pr(aX \leq x) = \Pr\left(X \geq \frac{x}{a}\right) &= \int_{\frac{x}{a}}^{\infty} p(t)\mathrm{d}t \\
&= \int_{x}^{-\infty} p\left(\frac{u}{a}\right)\frac{1}{a}\mathrm{d}u \qquad (u = ta) \\
&= -\int_{-\infty}^{x} p\left(\frac{u}{a}\right)\frac{1}{a}\mathrm{d}u
\end{aligned}
$$

より $-\frac{1}{a}p\left(\frac{x}{a}\right)$ である。よって一般の $a \in \mathbb{R} \setminus \{0\}$ について aX の確率密度関数は $\frac{1}{|a|}p\left(\frac{x}{a}\right)$ である。

$X + Y$ の確率密度関数

独立確率変数 X と Y がそれぞれ確率密度関数 p_X と p_Y を持つとする。
このとき、$X + Y$ の確率密度関数は

$$
\begin{aligned}
\Pr(X + Y \leq z) &= \int_{x+y \leq z} p_X(x) p_Y(y) \mathrm{d}x \mathrm{d}y \\
&= \int_{-\infty}^{\infty} p_Y(y) \left(\int_{-\infty}^{z-y} p_X(x) \mathrm{d}x \right) \mathrm{d}y \qquad (\text{多重積分} \to \text{逐次積分}) \\
&= \int_{-\infty}^{\infty} p_Y(y) \left(\int_{-\infty}^{z} p_X(x' - y) \mathrm{d}x' \right) \mathrm{d}y \qquad (x' = x + y) \\
&= \int_{-\infty}^{z} \left(\int_{-\infty}^{\infty} p_Y(y) p_X(x' - y) \mathrm{d}y \right) \mathrm{d}x'
\end{aligned}
$$

より $(p_X * p_Y)(x) := \int_{-\infty}^{\infty} p_X(x - y) p_Y(y) \mathrm{d}y$ である。これを確率密度関数 p_X と p_Y の畳み込みという。

正規分布の密度関数

$Z \sim N(\mu, \sigma^2)$ のとき、$Z + a$ と aZ の確率密度関数は以下のようになる。

$$p(x - a) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-a-\mu)^2}{2\sigma^2}}$$

$$\frac{1}{|a|} p\left(\frac{x}{a}\right) = \frac{1}{\sqrt{2\pi a^2\sigma^2}} \mathrm{e}^{-\frac{(x/a-\mu)^2}{2\sigma^2}}$$

$$= \frac{1}{\sqrt{2\pi a^2\sigma^2}} \mathrm{e}^{-\frac{(x-a\mu)^2}{2a^2\sigma^2}}$$

正規分布の密度関数の畳み込み

$$\int_{-\infty}^{\infty} p_{X_2}(y)p_{X_1}(x-y)\mathrm{d}y = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_2^2}}\mathrm{e}^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}} \frac{1}{\sqrt{2\pi\sigma_1^2}}\mathrm{e}^{-\frac{(x-y-\mu_1)^2}{2\sigma_1^2}}\mathrm{d}y$$

$$= \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2}}\mathrm{e}^{-\frac{\sigma_1^2(y-\mu_2)^2+\sigma_2^2(x-y-\mu_1)^2}{2\sigma_1^2\sigma_2^2}}\mathrm{d}y$$

$$= \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2}}\mathrm{e}^{-\frac{(\sigma_1^2+\sigma_2^2)\left(y-\frac{(\mu_1-x)\sigma_2^2+\mu_2\sigma_1^2}{\sigma_1^2+\sigma_2^2}\right)^2+\sigma_1^2\mu_2^2+\sigma_2^2(x-\mu_1)^2-\frac{((\mu_1-x)\sigma_2^2+\mu_2\sigma_1^2)^2}{\sigma_1^2+\sigma_2^2}}{2\sigma_1^2\sigma_2^2}}\mathrm{d}y$$

$$= \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}\mathrm{e}^{-\frac{\sigma_1^2\mu_2^2+\sigma_2^2(x-\mu_1)^2-\frac{((\mu_1-x)\sigma_2^2+\mu_2\sigma_1^2)^2}{\sigma_1^2+\sigma_2^2}}{2\sigma_1^2\sigma_2^2}}$$

$$= \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}\mathrm{e}^{-\frac{(x-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}}$$

$N(\mu_1+\mu_2,\ \sigma_1^2+\sigma_2^2)$ の確率密度関数

特性関数を使った証明

任意の独立確率変数 X と Y と $a \in \mathbb{R}$ について

$$\begin{aligned}
\varphi_{X+a}(t) &= \mathbb{E}\left[e^{it(X+a)}\right] = \left(\mathbb{E}\left[e^{itX}\right] \cdot e^{ita}\right) = \varphi_X(t)e^{iat} \\
\varphi_{aX}(t) &= \mathbb{E}\left[e^{it(aX)}\right] = \varphi_X(at) \\
\varphi_{X+Y}(t) &= \mathbb{E}\left[e^{it(X+Y)}\right] = \mathbb{E}\left[e^{itX}\right]\mathbb{E}\left[e^{itY}\right] = \varphi_X(t)\varphi_Y(t)
\end{aligned}$$

$X \sim N(\mu, \sigma^2)$ について

$$\varphi_X(t) = e^{i\mu t - \frac{\sigma^2 t^2}{2}}$$

より、

$$\varphi_{X+a}(t) = e^{i(\mu+a)t - \frac{\sigma^2 t^2}{2}} \qquad (N(\mu + a, \sigma^2) \text{ の特性関数})$$

$$\varphi_{aX}(t) = e^{i\mu a t - \frac{a^2\sigma^2 t^2}{2}} \qquad (N(a\mu, a^2\sigma^2) \text{ の特性関数})$$

$$\varphi_{X_1+X_2}(t) = e^{i(\mu_1+\mu_2)t - \frac{(\sigma_1^2+\sigma_2^2)t^2}{2}} \qquad (N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2) \text{ の特性関数})$$

ポアソン分布の再生性

確率変数 X が平均 $\lambda > 0$ のポアソン分布に従うことを $X \sim \mathrm{Poisson}(\lambda)$ と表す。

Lemma

独立確率変数 X と Y が $X \sim \mathrm{Poisson}(\lambda_1)$、$Y \sim \mathrm{Poisson}(\lambda_2)$ のとき $X + Y \sim \mathrm{Poisson}(\lambda_1 + \lambda_2)$.

Proof.

$Z \sim \mathrm{Poisson}(\lambda)$ について $\varphi_Z(t) = \mathrm{e}^{\lambda(\mathrm{e}^{it}-1)}$ より

$$\begin{aligned}
\varphi_{X+Y}(t) &= \varphi_X(t)\varphi_Y(t) \\
&= \mathrm{e}^{\lambda_1(\mathrm{e}^{it}-1)}\mathrm{e}^{\lambda_2(\mathrm{e}^{it}-1)} \\
&= \mathrm{e}^{(\lambda_1+\lambda_2)(\mathrm{e}^{it}-1)}
\end{aligned}$$

であり、これは $\mathrm{Poisson}(\lambda_1 + \lambda_2)$ に従う確率変数の特性関数であるため $X + Y \sim \mathrm{Poisson}(\lambda_1 + \lambda_2)$. $\square$

密度関数

$J \subseteq \mathbb{R}$ を区間とし、$f : J \to \mathbb{R}$ を J の内点で微分可能で $f'(x) > 0$ とする。$x \in \mathrm{Image}(f)$ について

$$\begin{aligned} \Pr(f(X) \leq x) &= \Pr\left(X \leq f^{-1}(x)\right) \\ &= \int_{-\infty}^{f^{-1}(x)} p(z)\mathrm{d}z \\ &= \int_{-\infty}^{x} p\left(f^{-1}(u)\right) \frac{1}{f'(f^{-1}(u))}\mathrm{d}u \qquad (u = f(z)) \end{aligned}$$

よって $f(X)$ の確率密度関数は $\frac{1}{f'(f^{-1}(x))} p(f^{-1}(x))$ である。

課題

- 表が出る確率が p のコインを N 回投げて表が出る回数が区間 $\left[pN - a\sqrt{N},\ pN + a\sqrt{N}\right]$ に入る確率が $N \to \infty$ 極限で

$$\int_{-b}^{b} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}x^2} \mathrm{d}x$$

  に収束するような $b > 0$ を p と a を使って表せ。

- U を $[0,1]$ 上の一様分布とする (確率密度関数は $p(x) = \mathbb{1}_{\{x \in [0,1]\}}$ である)。このとき、$-2\log U$ の確率密度関数をもとめよ。

- $Z \sim N(0,1)$ のとき、Z^2 の確率密度関数をもとめよ。